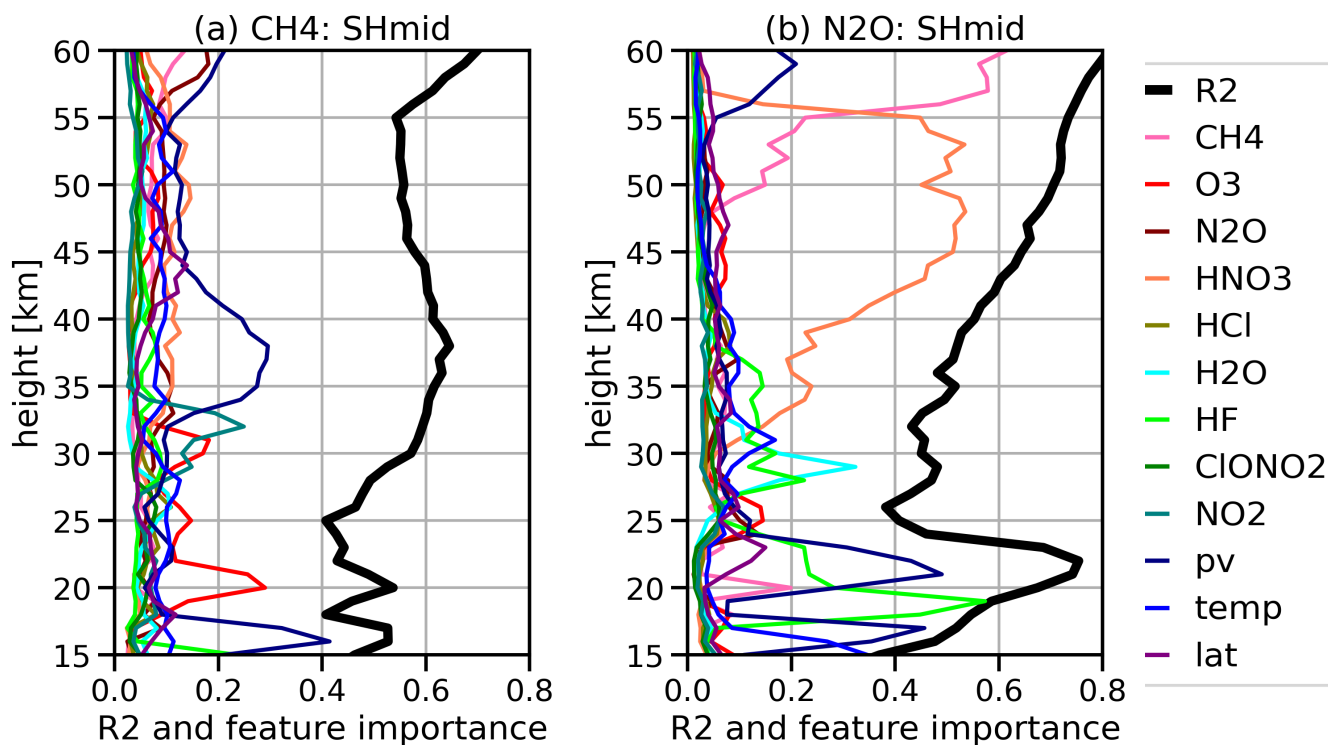


# **Supplementary material: Using machine-learning to construct TOMCAT model and occultation measurement-based stratospheric methane (TCOM-CH<sub>4</sub>) and nitrous oxide (TCOM-N<sub>2</sub>O) profile data sets**

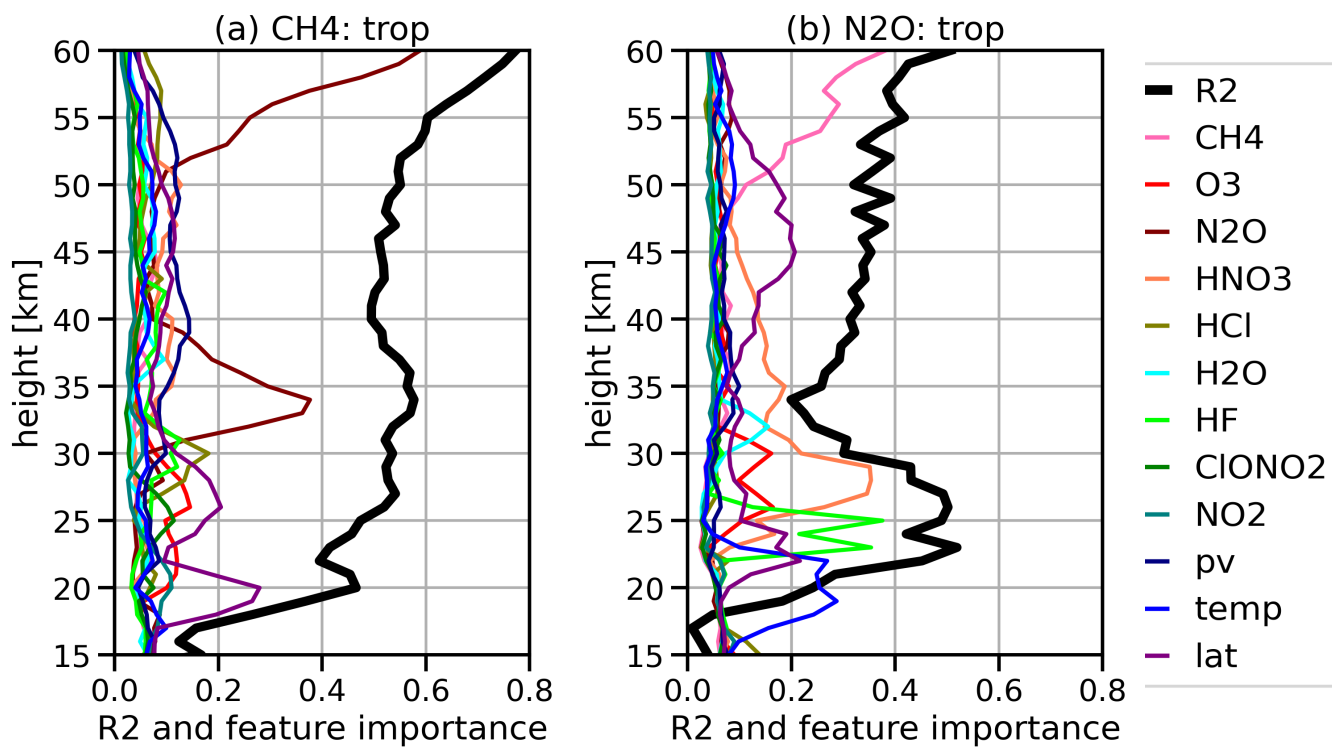
Sandip S. Dhomse<sup>1,2</sup> and Martyn P. Chipperfield<sup>1,2</sup>

<sup>1</sup>School of Earth and Environment, University of Leeds, Leeds, UK

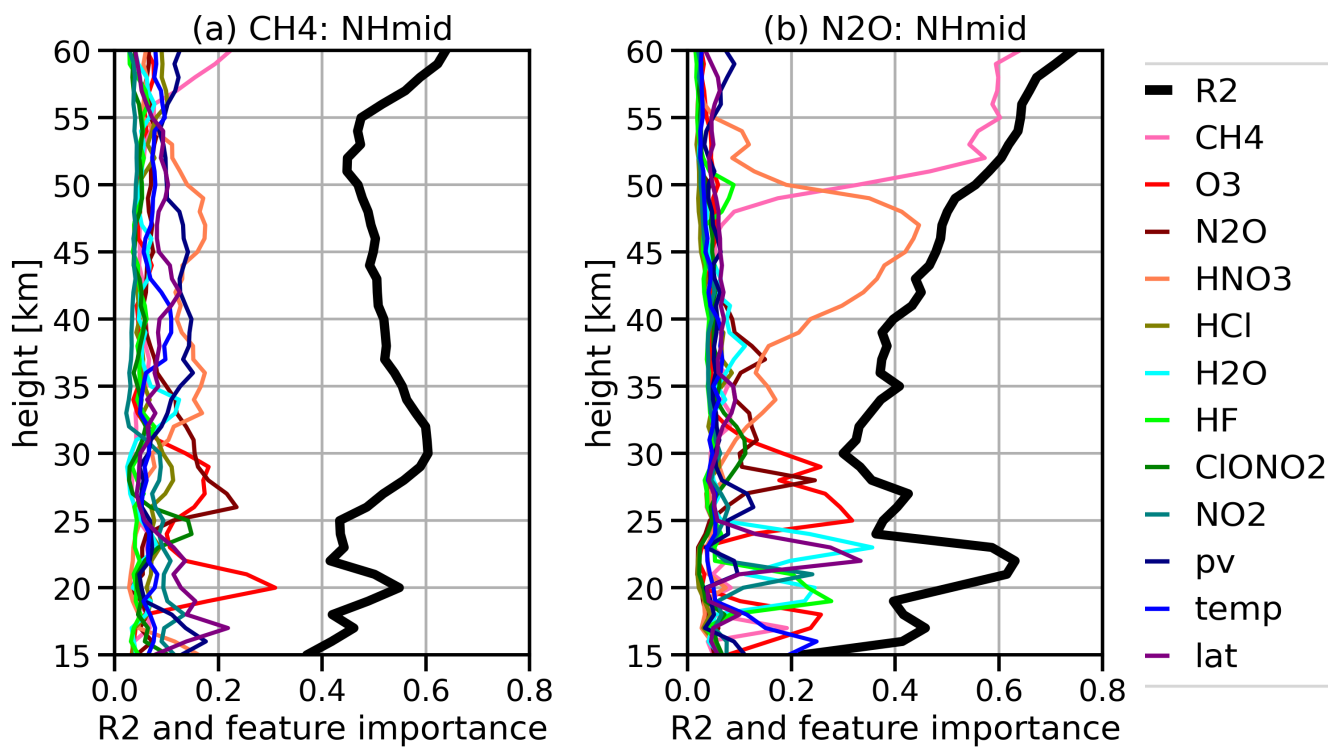
<sup>2</sup>National Centre for Earth Observation, University of Leeds, Leeds, UK



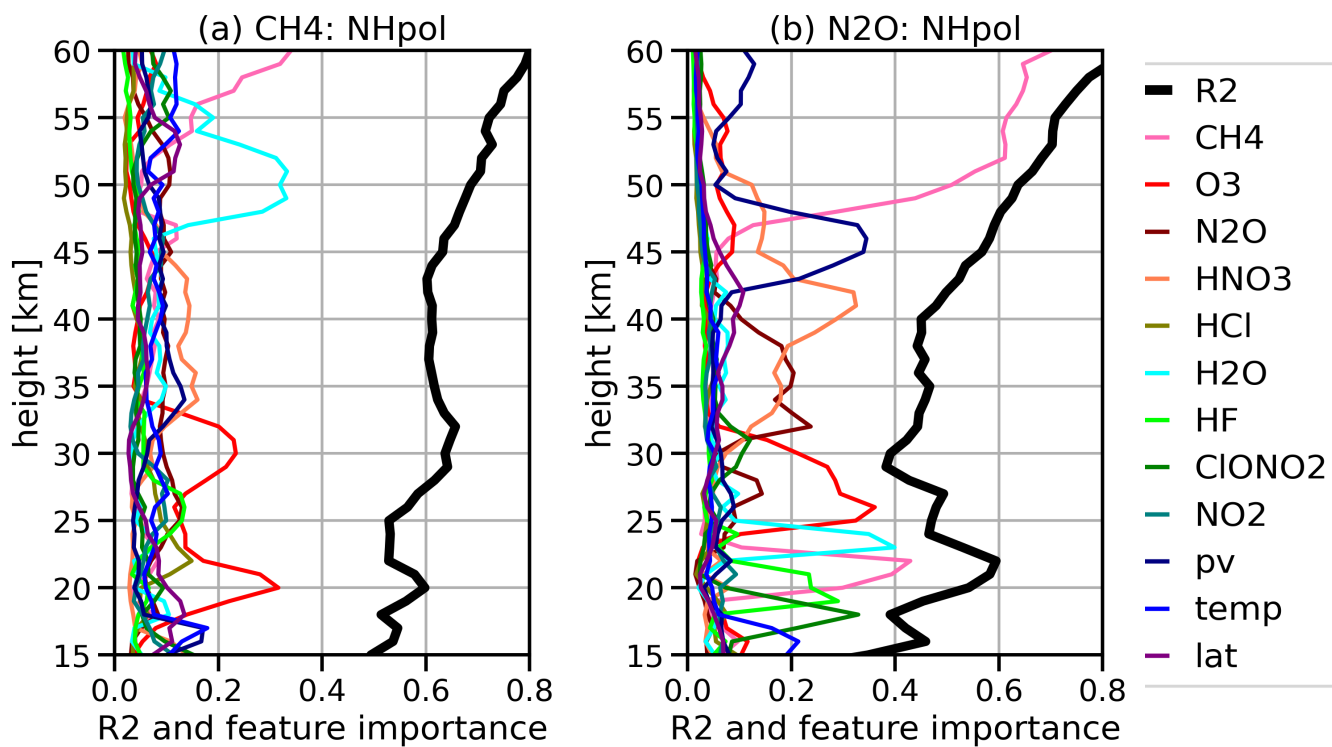
**Figure S1.** Vertical profiles of the variance ( $R^2$ ) and feature importances estimated by XGBoost regression models for the TOMCAT-observation differences for (a) CH<sub>4</sub> (1991-2018) and (b) N<sub>2</sub>O (2004-2018, ACE only) for the South Hemisphere mid-latitude (SHmid, 20°S–70°S) latitude bin. See equation 1 and subsequent information in the manuscript about the features (total 13) or variables used in the XGBoost regression model.



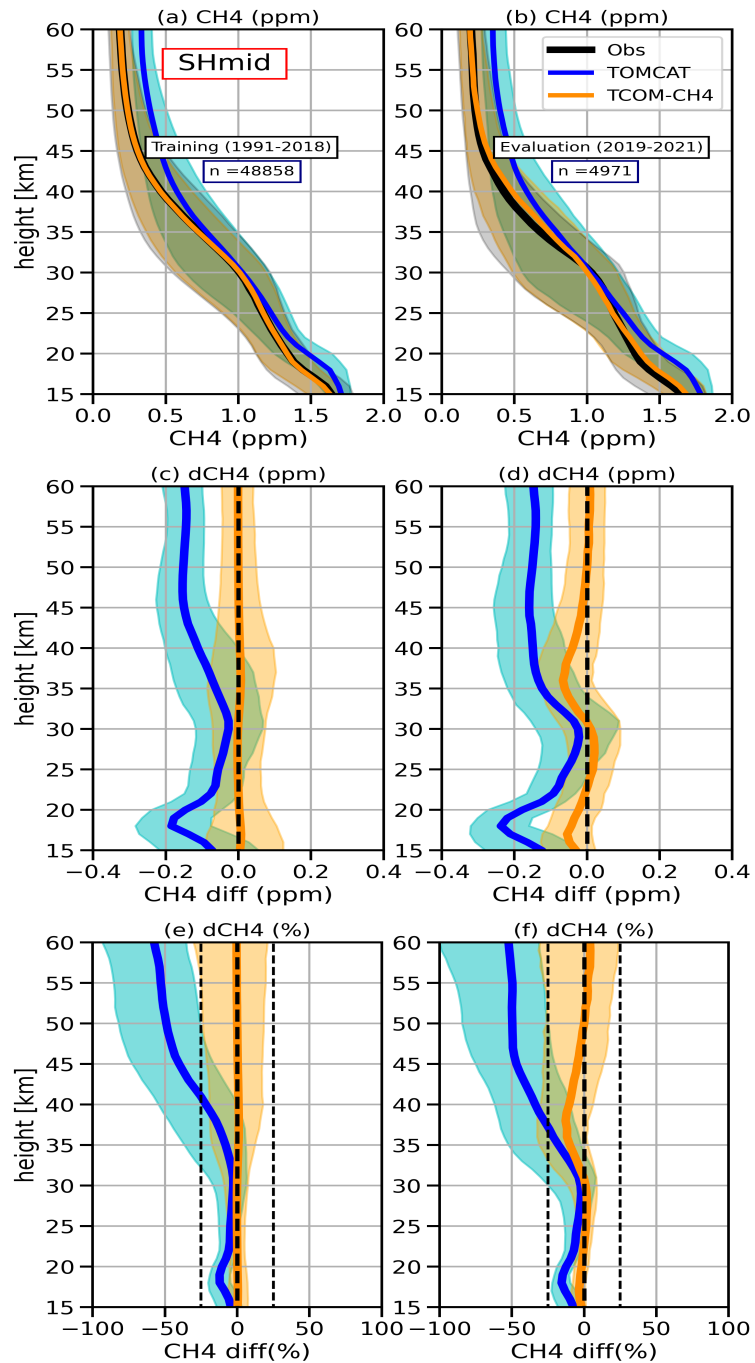
**Figure S2.** Same as S1, but for tropical latitude band (40° S–40° N)



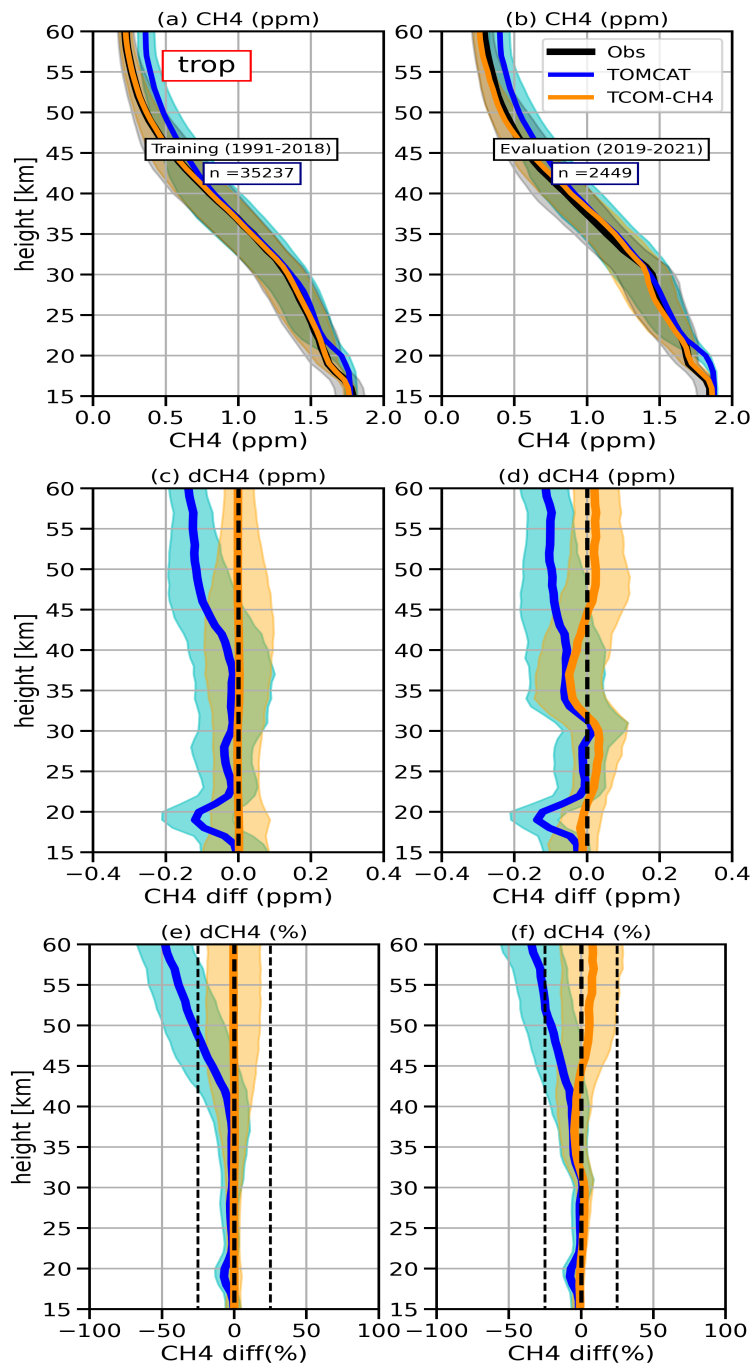
**Figure S3.** Same as S1, but for northern hemisphere mid-latitude (NHmid) band (20°N–70°N)



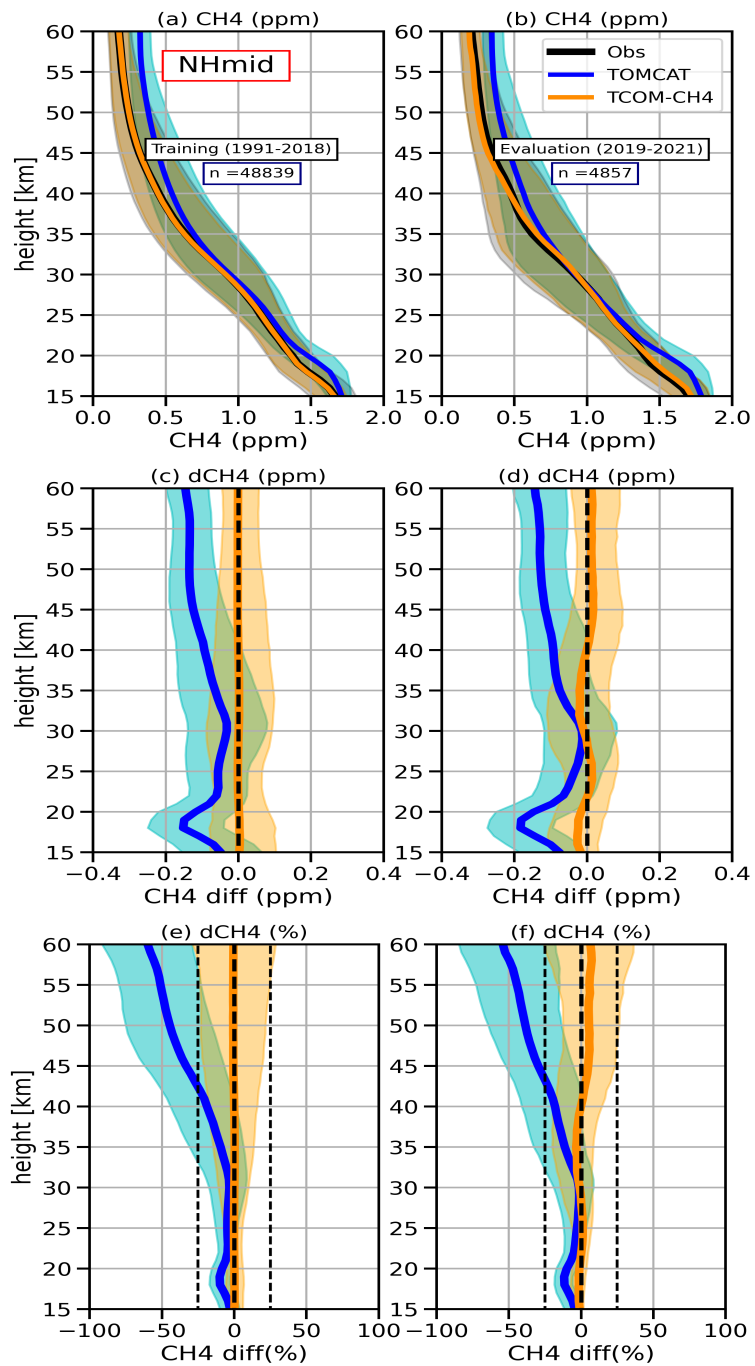
**Figure S4.** Same as S1, but for northern hemisphere polar (NHpol) latitude band (50°N–90°N)



**Figure S5.** Panels (a) and (b). Comparison between TOMCAT (blue), TCOM-CH4 (orange) and satellite measurement-based (black) CH<sub>4</sub> profiles for SHmid (20°S–70°S) latitude band. Solid lines indicate median profiles while shaded regions show 10th and 90th percentile range. Comparisons are shown for training (1992-2018) and evaluation (2019-2021) periods in panels (a, left) and (b, right), respectively. Panels (c) - (f). Differences between TOMCAT and TCOM-CH4 w. r. t. satellite data sets in absolute units (ppm) and percent. Right (c and e) and left (d and f) panels show differences for the training (1992-2018) and evaluation (2019-2021) periods.

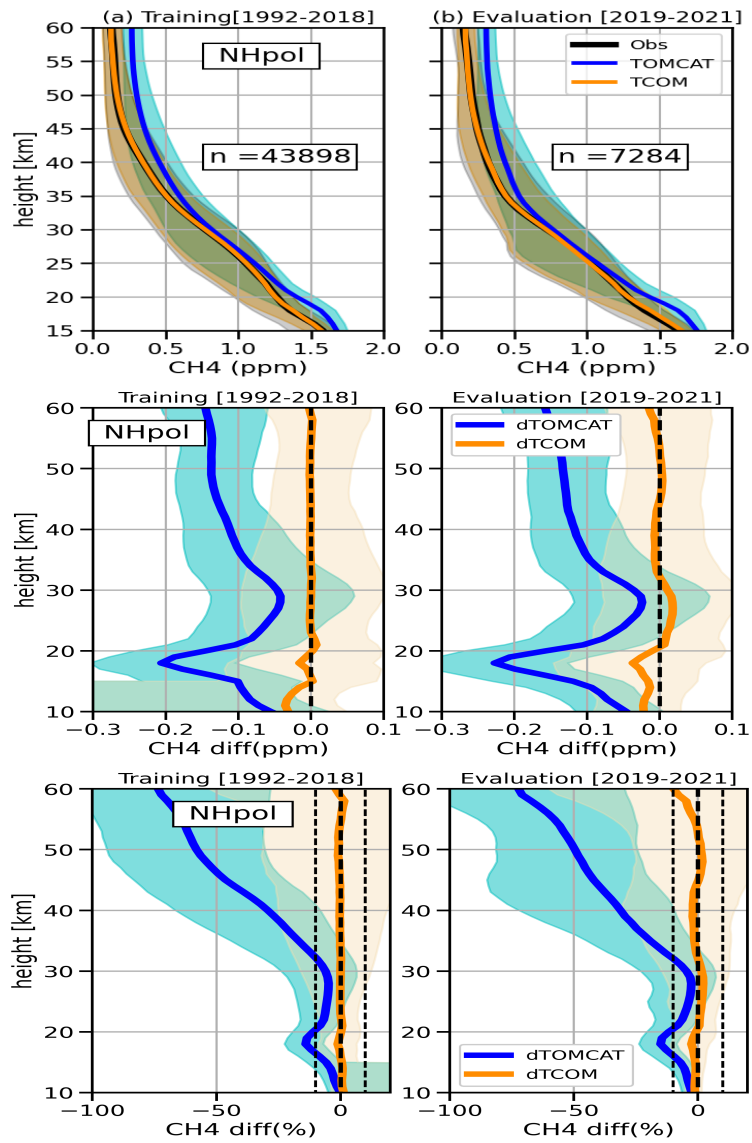


**Figure S6.** Same as S5, but for tropical (trop) latitude band (40°S–40°N)

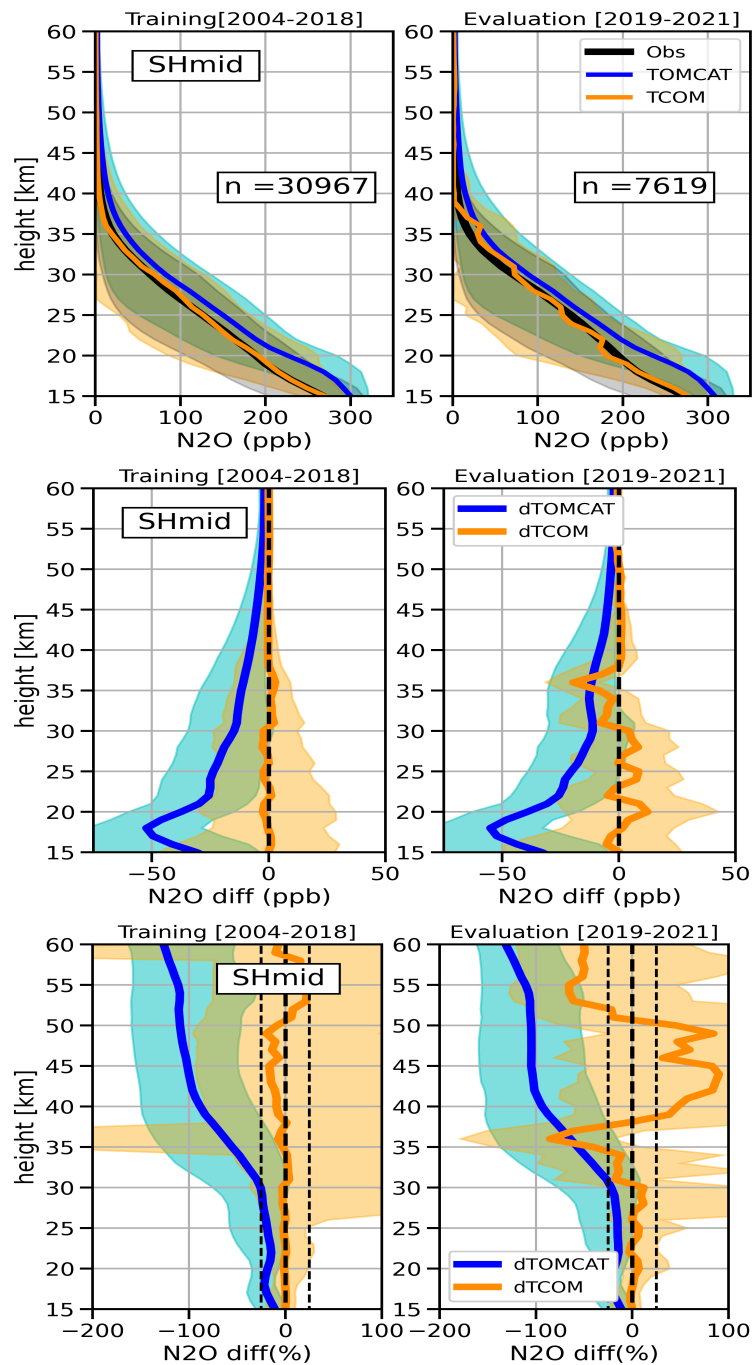


**Figure S7.** Same as S5, but for northern hemisphere mid-latitude (NHmid) band (20°N–70°N)

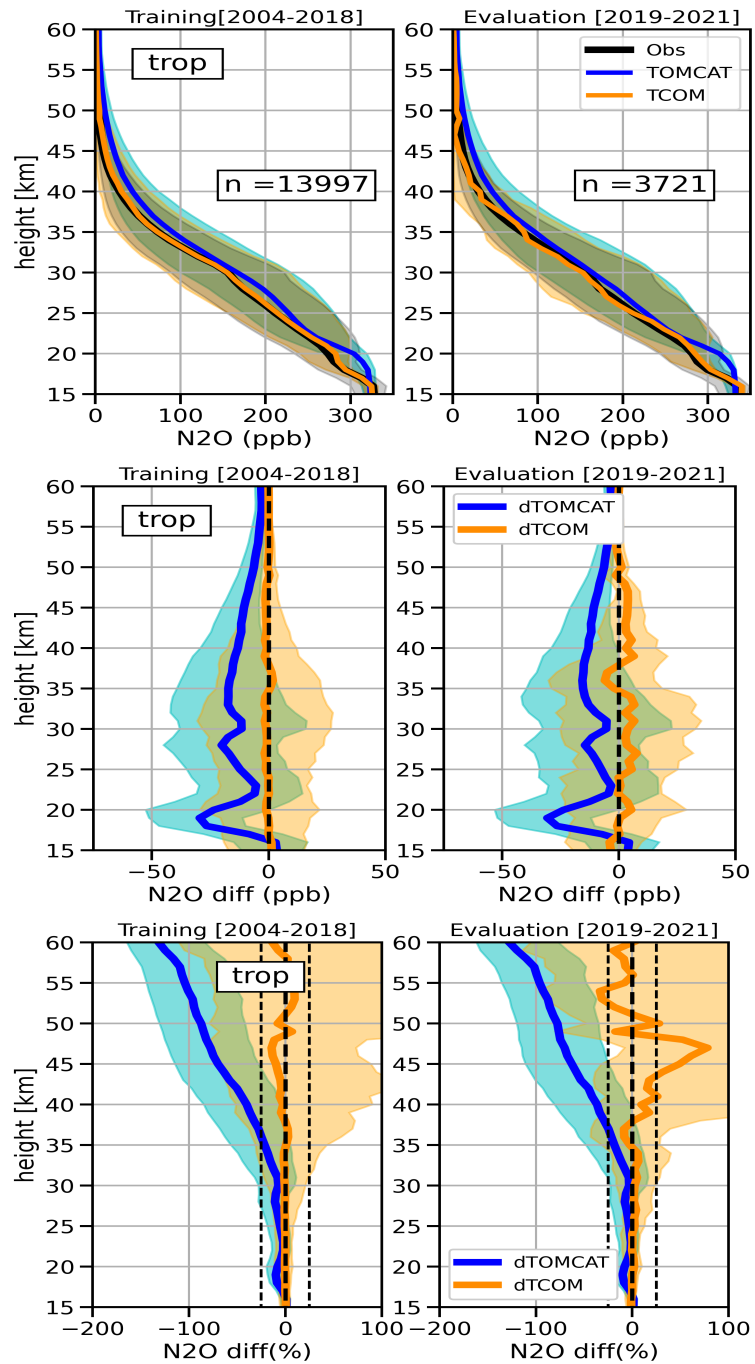




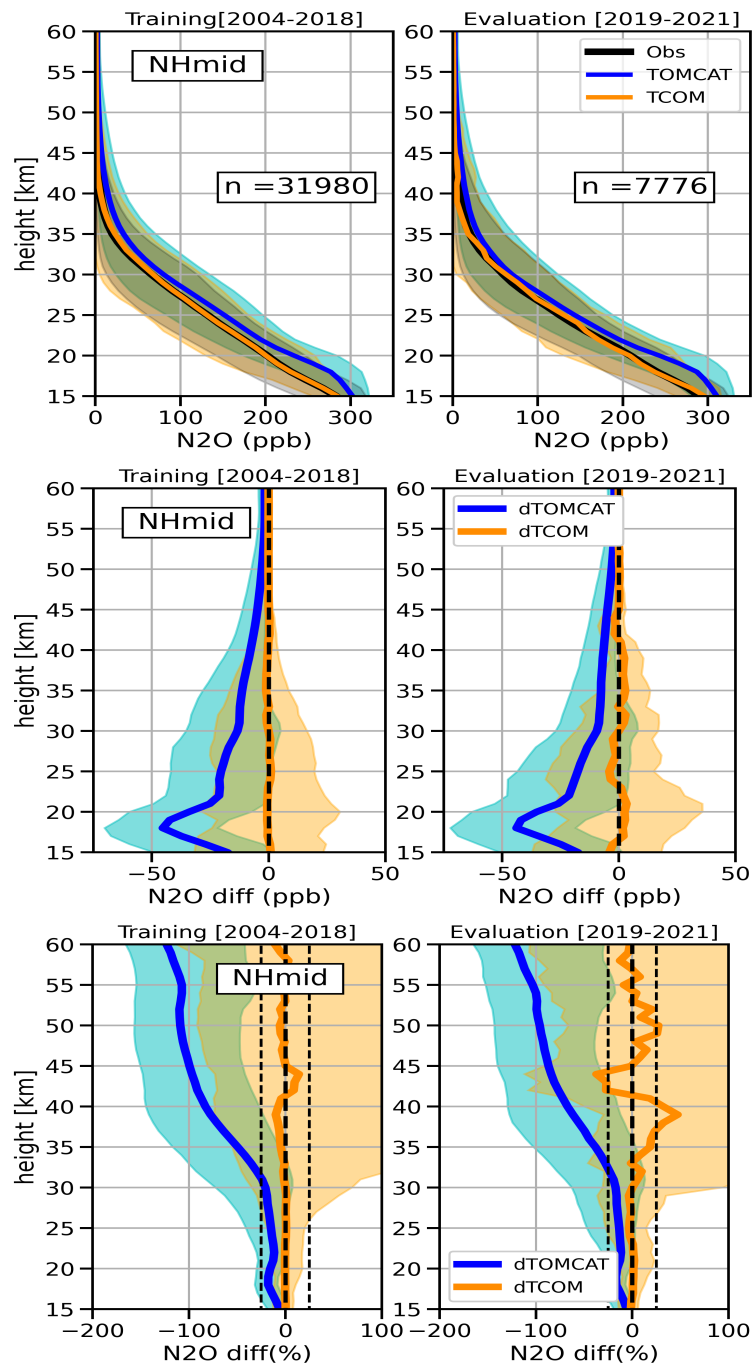
**Figure S8.** Same as S5, but for northern hemisphere polar-latitude (NHpol) band (50°N-90°N)



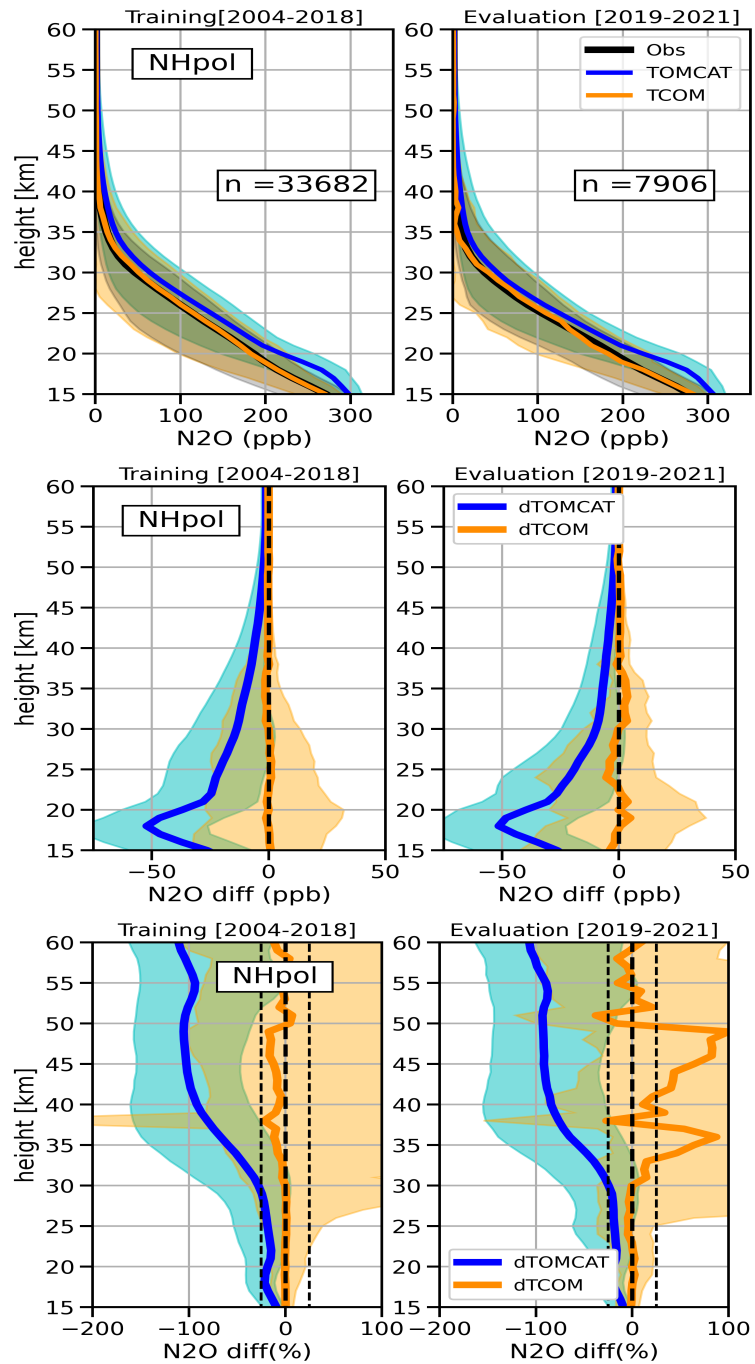
**Figure S9.** Panels (a) and (b). Comparison between TOMCAT (blue), TCOM-N<sub>2</sub>O (orange) and ACE-FTS satellite (black) N<sub>2</sub>O profiles measurements for SHmid (20°S–70°S) latitude band. Solid lines indicate median profiles while shaded regions show 10th and 90th percentile range. Comparisons are shown for training (1992-2018) and evaluation (2019-2021) periods in panels (a, left) and (b, right), respectively. Panels (c) - (f). Differences between TOMCAT and TCOM-N<sub>2</sub>O w. r. t. satellite data sets in absolute units (ppm) and percent. Right (c and e) and left (d and f) panels show differences for the training (1992-2018) and evaluation (2019-2021) periods.



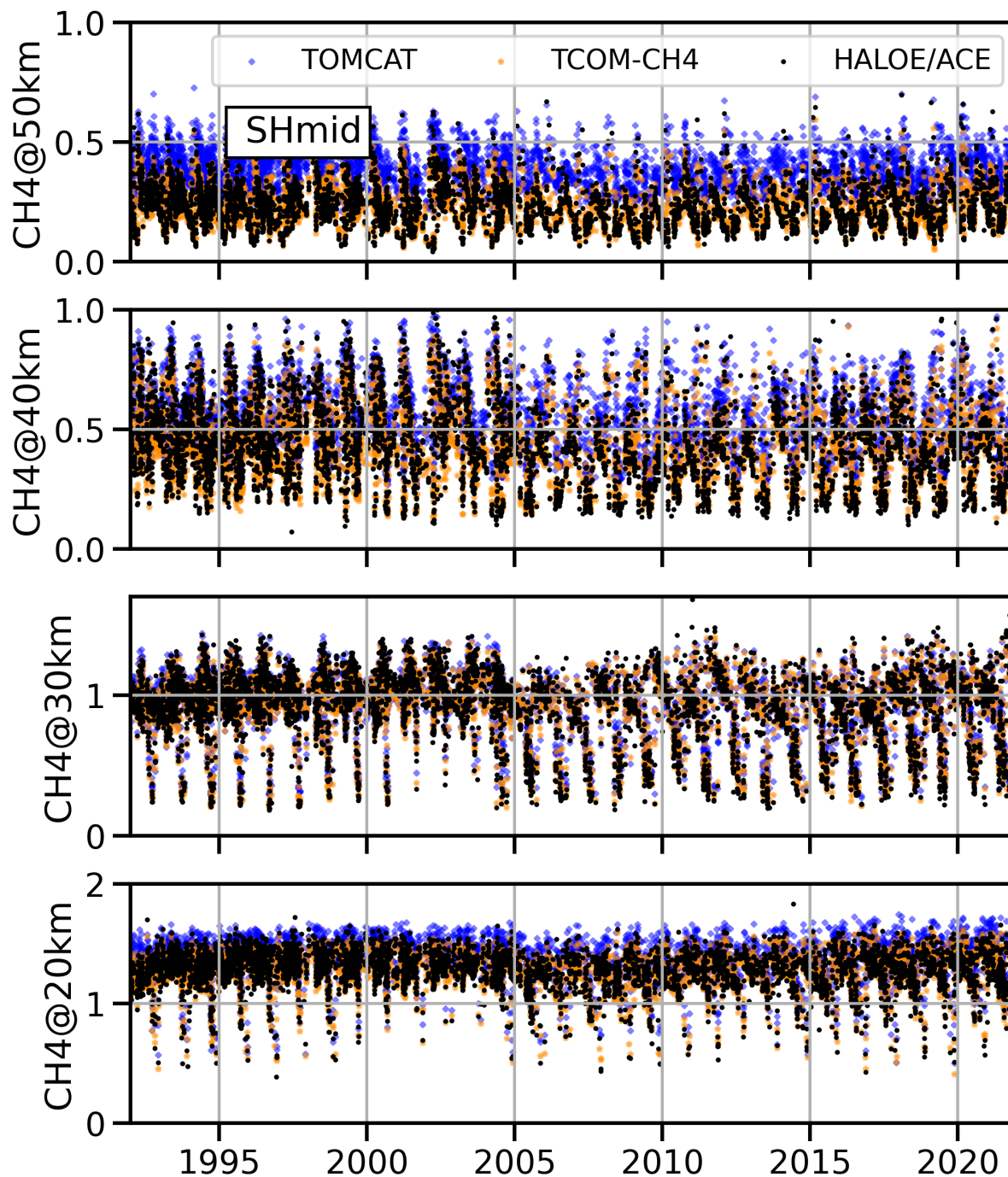
**Figure S10.** Same as S9, but for tropical (trop) latitude band (40°S–40°N)



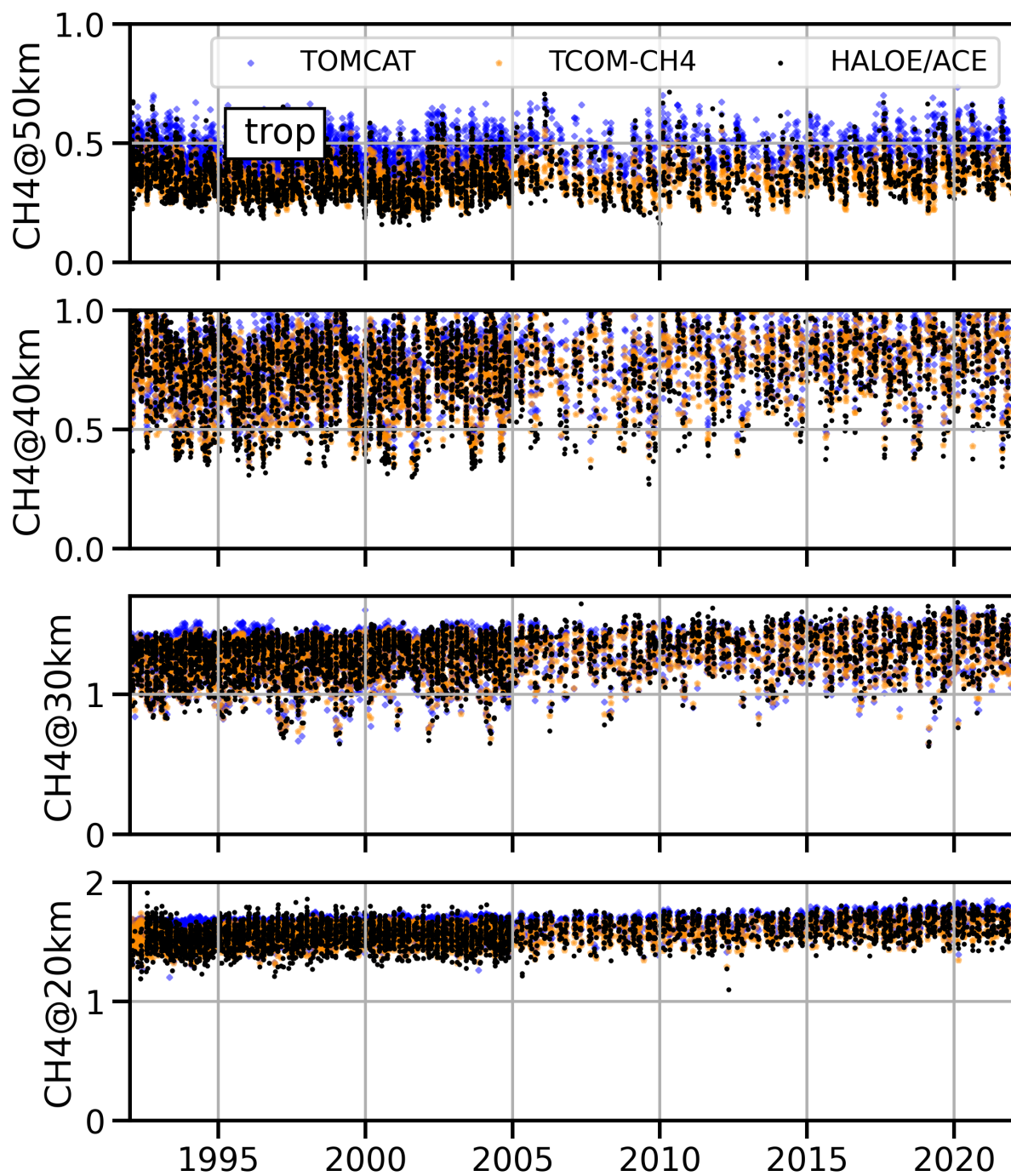
**Figure S11.** Same as S9, but for northern hemisphere mid-latitude (NHmid) band (20°N–70°N)



**Figure S12.** Same as S9, but for northern hemisphere polar-latitude (NHpol) band (50°N–90°N)



**Figure S13.** Time evolution (1992–2021) of CH<sub>4</sub> from TOMCAT (blue crosses), TCOM-CH<sub>4</sub> (orange diamonds) and satellite data (black dots) for SHmid (20°S–70°S) at 20, 30, 40 and 50 km. Note that for clarity only 10% (every 10th) of data points are shown and due to sharp gradient in vertical distribution, the y axis range varies between the panels.



**Figure S14.** Same as S13, but for tropical (trop) latitude band (40°S–40°N)

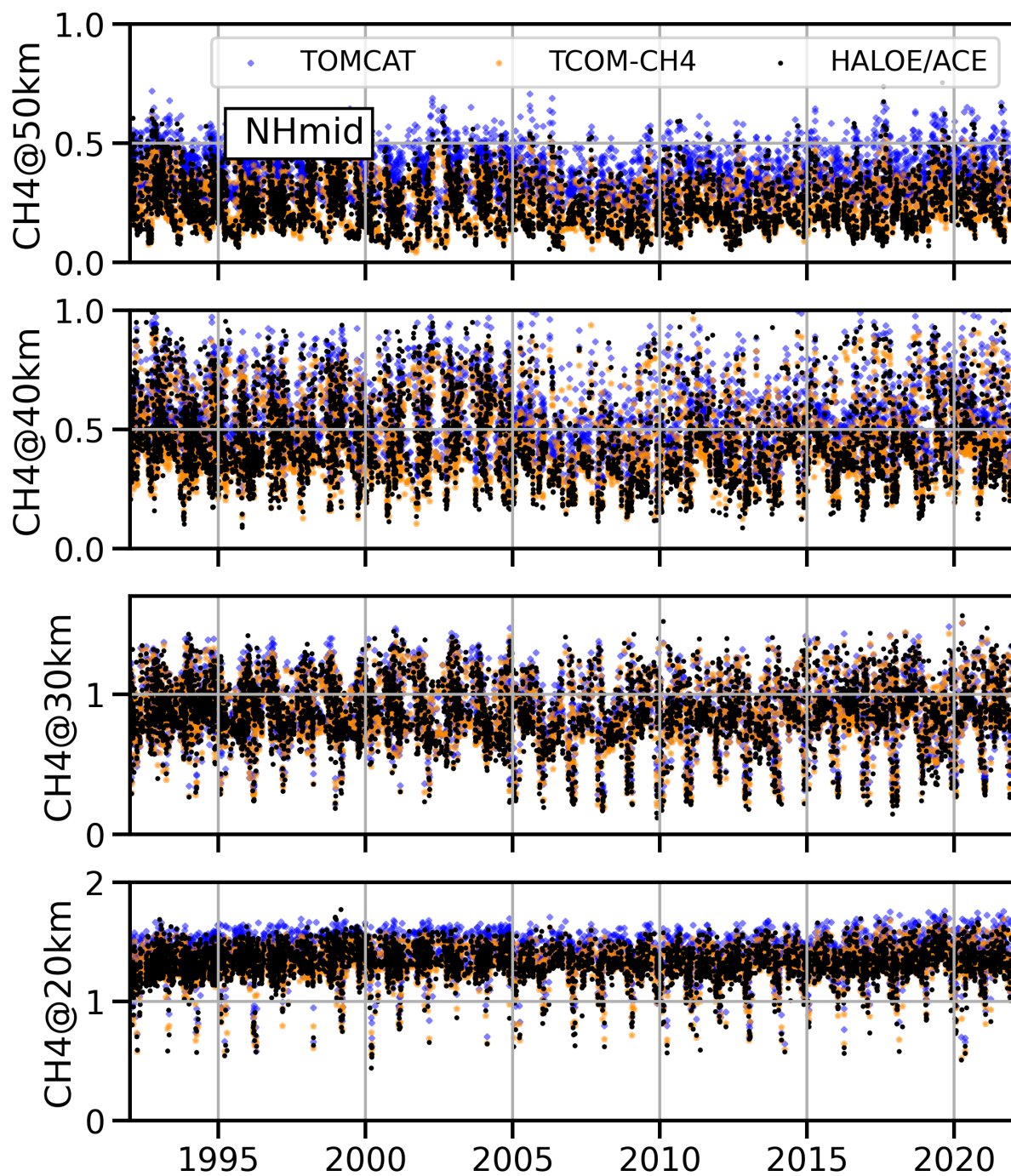
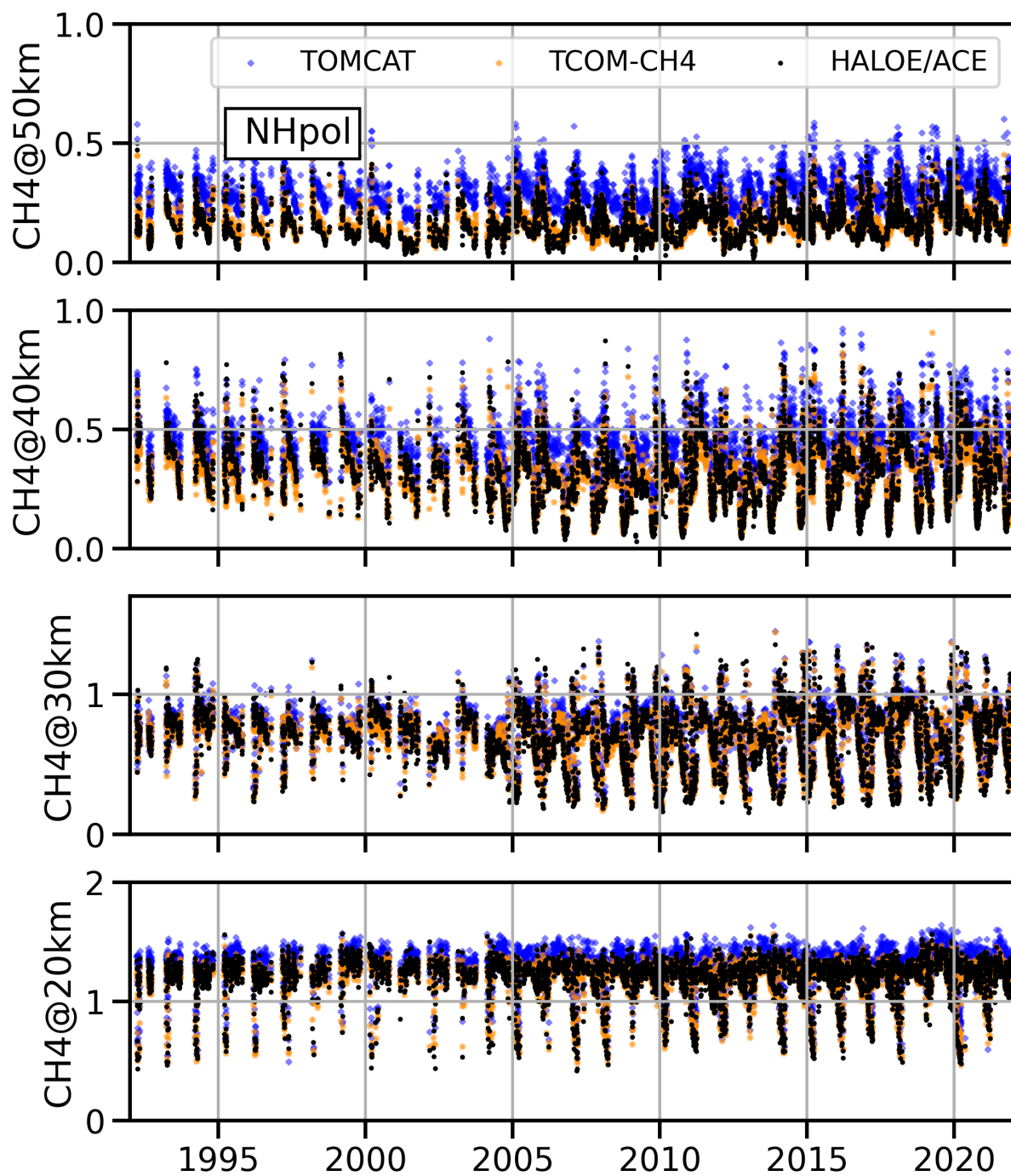
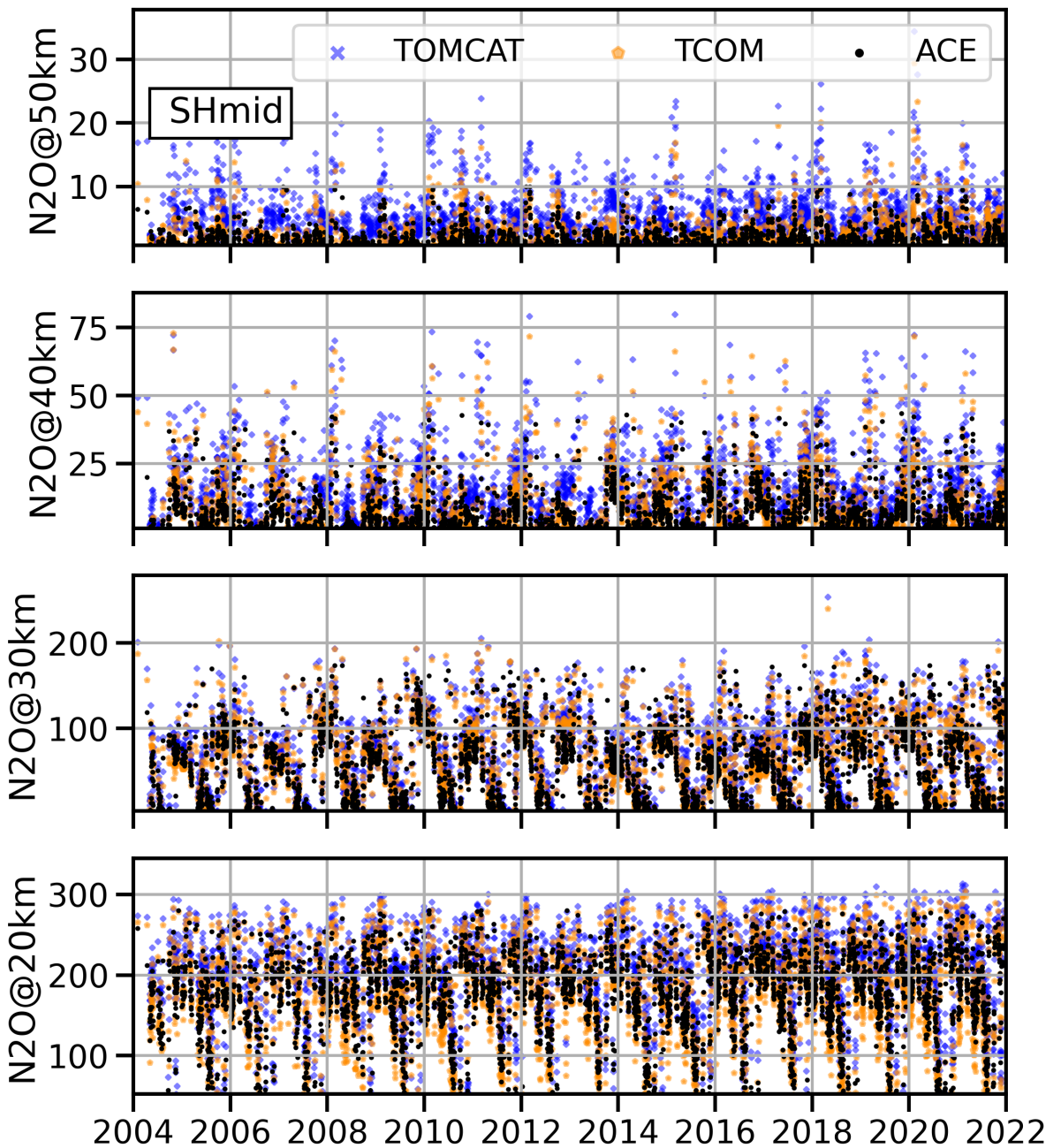


Figure S15. Same as S13, but for northern hemisphere mid-latitude (NHmid) band (20°N–70°N)

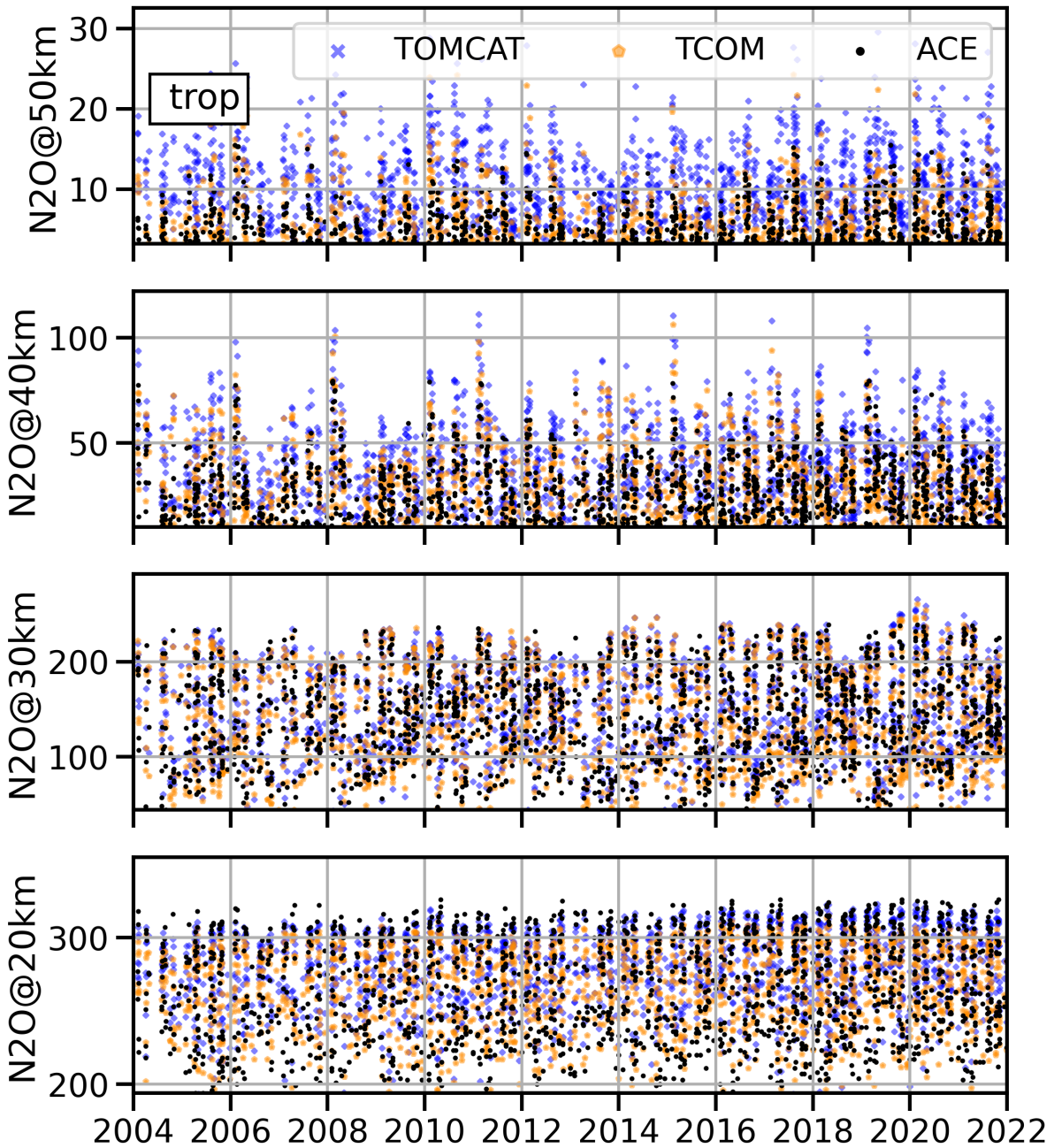




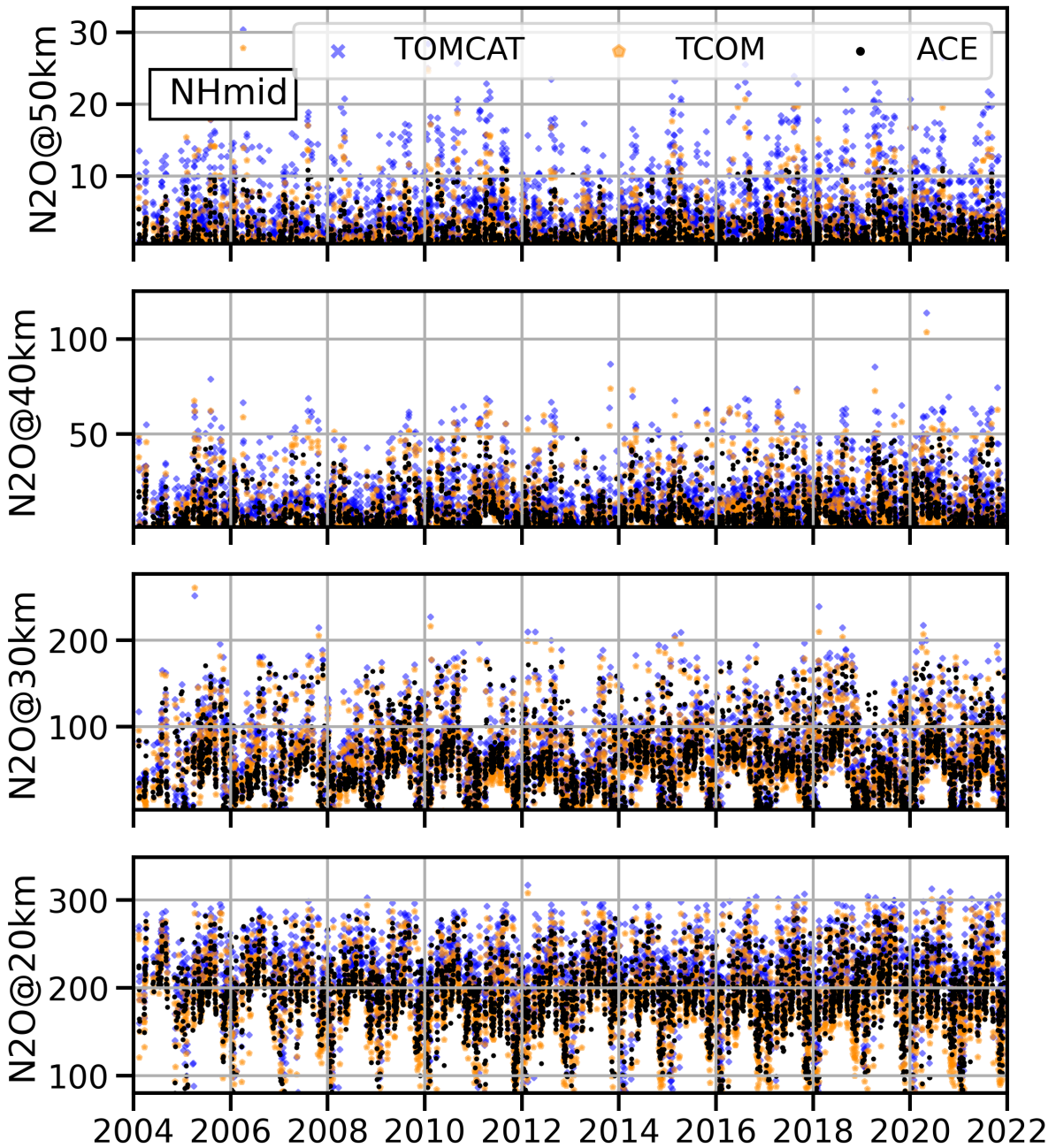
**Figure S16.** Same as S13, but for northern hemisphere polar-latitude (NHpol) band (50°N–90°N)



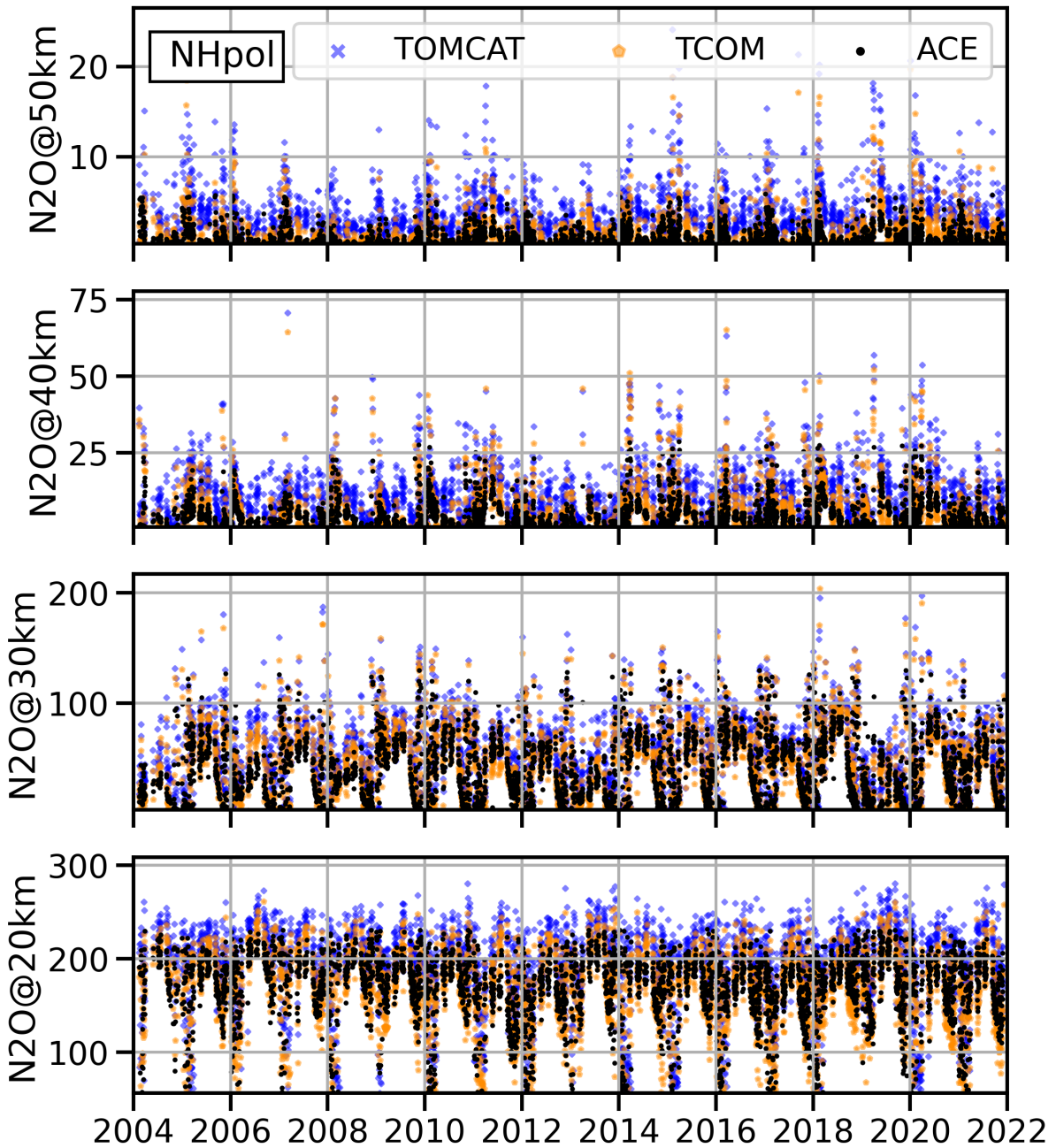
**Figure S17.** Time evolution (1992–2021) of  $\text{CH}_4$  from TOMCAT (blue crosses), TCOM- $\text{CH}_4$  (orange diamonds) and satellite data (black dots) for SHmid ( $20^\circ\text{S}$ – $70^\circ\text{S}$ ) at 20, 30, 40 and 50 km. Note that for clarity only 10% (every 10th) of data points are shown and due to sharp gradient in vertical distribution, the y axis range varies between the panels.



**Figure S18.** Same as S17, but for tropical (trop) latitude band (40°S–40°N)



**Figure S19.** Same as S17, but for northern hemisphere mid-latitude (NHmid) band (20°N–70°N)



**Figure S20.** Same as S17, but for northern hemisphere polar-latitude (NHpol) band (50°N–90°N)